JOURNAL OF APPROXIMATION THEORY 16, 329–337 (1976)

Adaptive Approximation

JOHN R. RICE*

Division of Mathematical Sciences, Purdue University, West Lafayette, Indiana 47907

Communicated by Oved Shisha

Received July 19, 1974

I. INTRODUCTION

The purpose of this paper is to describe a general class of adaptive methods for determining approximations. These methods are primarily of interest for piecewise polynomial approximation in L_p norms ($1 \leqslant p \leqslant \infty$), but are not necessarily so restricted. They are applicable to a broad class of functions including any that are likely to occur in practice. We analyze the rate of convergence of these methods and show that they have maximum degree of convergence.

We mention concrete realizations of these methods, which allow one to compute *smooth* approximations rapidly. Fast methods for unsmooth (e.g., only continuous) approximations have been known for some time although they are not described in the literature and their convergence properties have not been analyzed formally.

We also point out that these results provide simpler and more constructive proofs of some earlier results on the degree of convergence for nonlinear spline approximation.

Finally, for the sake of completeness, we note that the term adaptive approximation (for piecewise polynomials) has been used for a completely different approach by Dodson [4].

II. PRELIMINARIES

An adaptive approximation method involves a local approximation operator T_I, which associates with $f(x)$ an approximation $A_I(f, x)$ on the interval I, i.e.,

$$T_I : f(x) \to A_I(f, x).$$

* This work was supported in part by Grant GP 32940X from the National Science Foundation.

A simple example of this is linear interpolation at the end points of the interval I. Associated with the adaptive method is a tolerance $\epsilon > 0$ and an interval I is *active* if

$$\|f(x) - A_I(f, x)\|_I = \|(1 - T_I)f\|_I \geqslant \epsilon,$$

otherwise it is *discarded*. The subscript on the norm indicates restriction to the interval I. The number ϵ is not necessarily the desired approximation accuracy.

Adaptive Approximation Method

1. Initially we are given $f(x)$, the interval [0, 1], a tolerance $\epsilon > 0$, a local approximation operator T_I and a norm $\| \ \|$. The collection of active intervals initially consists of [0, 1].

2. Choose an active interval I and subdivide it into two parts IL and IR. Find $T_{IL}f$ and $T_{IR}f$.

3. Discard IL or IR if $\|(1 - T_{IL})f\|_{IL} < \epsilon$ or $\|(1 - T_{IR})f\|_{IR} < \epsilon$, otherwise return them to the collection of active intervals.

4. Return to step 2 as long as there are any active intervals.

When the adaptive method terminates, we know that the local error of approximation on each interval is less than ϵ and this allows one to estimate the global error depending on the nature of the norm used. The global approximation is, of course, just the collection of local approximations $A_I(f, x)$. For simplicity, we assume that each interval is halved and thus each interval is of the form $[x, x + 2^{-k}]$ for some value of k and we may represent it by the pair (x, k). For specific local approximation operators and suitable functions $f(x)$ we have a bound on the error $\|(1 - T_I)f\|_I$ and we denote this by ERROR(x, k). In actual use of an adaptive method the decision to discard an interval is based on ERROR(x, k) rather than the exact value of $\|(1 - T_I)f\|_I$, which may be difficult (or impossible) to compute. In a typical situation we would assume that $f(x)$ belongs to some smoothness class parameterized by a number n (say piecewise in C^n) and then we would know that

$$\text{ERROR}(x, k) \leqslant K(x)(2^{-k})^r,$$

where r is simply related to n (say $r = n$, or $n + 1$) and $K(x)$ is perhaps the value of some derivative of $f(x)$, or $\|f^{(n)}(x)\|_I$. The objective of this paper is to analyze the rate of convergence of adaptive approximation methods in such situations.

The convergence proofs draw heavily on results about an interval partition algorithm previously established in [7]. This algorithm is as follows:

PARTITION ALGORITHM.

1. *Initialization. We are given*

A. *Numbers* γ, $\beta < 1$ *and* $\epsilon > 0$.

B. *An empty set* M' *and a set* M *of intervals* I *with associated numbers* $\eta(I)$. M *contains a distinguished interval* I^*.

C. *A process* $P: I \to (IL, IR)$ *that divides an interval* I *into left and right subintervals such that*

(i) *If* $I = I^*$ *then* $\eta(IL) = \eta(IR) = \beta * \eta(I)$ *and* $I^* \leftarrow IL$ *or* $I^* \leftarrow IR$.

(ii) *If* $I \neq I^*$ *then* $\eta(IL) = \eta(IR) = \gamma * \eta(I)$.

2. *Operation. For* $I \in M$ *do*

$P: I \to (IL, IR)$

If $(\eta(IL) < \epsilon)$ *then* $IL \in M'$ *else* $IL \in M$

If $(\eta(IR) < \epsilon)$ *then* $IR \in M'$ *else* $IR \in M$.

The following results are established in [7].

THEOREM 1. *Consider the Partition Algorithm with* β, M, *and* $\eta(I)$ *for* $I \in M$ *specified. Let* $F(\gamma, \epsilon)$ *be the size of* M' *when the algorithm terminates, and then we have*

$$F(\gamma, \epsilon) = \mathcal{O}(\epsilon^{1/\log_2 \gamma}).$$

COROLLARY 1. *If the Partition Algorithm is modified so that each interval is divided into* m *parts or less and if there are* k *distinguished intervals, then the conclusion of Theorem* 1 *becomes*

$$F(\gamma, \epsilon) = \mathcal{O}(\epsilon^{1/\log_m \gamma}).$$

COROLLARY 2. *Consider a real valued function* g *defined on intervals with the property that* $I_1 \subseteq I_2$ *implies* $g(I_1) \leqslant g(I_2)$. *Suppose that in the interval division process* P *the factors* γ *and* β *are replaced by* $\gamma * g(IR)$, $\gamma * g(IL)$, $\beta * g(IR)$, *and* $\beta * g(IL)$ *as appropriate. Then the conclusion of Theorem* 1 *remains valid.*

III. TCHEBYCHEFF APPROXIMATION

We use the norm

$$\|f(x)\|_I = \max_{x \in I} |f(x)| = \|f(x)\|_{I,\infty},$$

and consider the functions $f(x)$ which satisfy the following assumption:

ASSUMPTION 1. *Assume $f(x)$ has singularities*

$$S = \{s_i \mid i = 1, 2, \ldots, R < \infty\},$$

and set

$$w(x) = \prod_{i=1}^{R} (x - s_i).$$

(i) *If $x_0 \notin S$ then $f^{(n)}(x)$ is continuous in a neighborhood of x_0 with $n \geq 1$.*

(ii) *There are constants K and α so that*

$$|f^{(n)}(x)| \leq K \mid w(x)|^{\alpha - n}.$$

This assumption states that $f(x)$ has n continuous derivatives except for a finite number R of algebraic singularities.

We consider local approximation operators T_I which satisfy:

ASSUMPTION 2. *Let s denote a point of singularity of $f(x)$ and set*

$$F_n(x, k) = \max_{t \in [x, x+2^{-k}]} |f^{(n)}(t)|.$$

There are constants n, K, and α (the same as in Assumption 1) so that:

(i) $\text{ERROR}(x, k) \leq KF_n(x, k)\, 2^{-kn}$, *if $[x, x + 2^{-k}]$ contains no singularity.*

(ii) $\text{ERROR}(x, k) \leq K2^{-k}$, *if $s \in [x, x + 2^{-k}]$.*

Note that the second part of this assumption implies that most common local approximation operators must be modified for intervals containing singularities. A little thought shows that even very crude modifications suffice to satisfy this part of the assumption.

THEOREM 2. *Assume $f(x)$ satisfies Assumption 1 with $\alpha > 0$. Consider an adaptive algorithm whose local operator satisfies Assumption 2. Then the global approximation $A(x)$ obtained when the algorithm terminates satisfies*

$$\|f(x) - A(x)\|_{[0,1],\infty} = \mathcal{O}(1/N^n),$$

where N is the number of pieces comprising $A(x)$.

Proof. The interval collection is initialized with $[0, 1]$ and the distinguished intervals are those that contain a singularity. One singularity may produce two distinguished intervals in case it is an end point of a subdivision, but clearly, there are at most $2R$ distinguished intervals. The numbers associated with the intervals are governed by Corollary 2 where

$g[x, x + 2^{-k}]$ is $F_n(x, k)$ as defined in Assumption 2. The values of γ and β of the partition algorithm are

$$\gamma = 2^{-n}, \qquad \beta = 2^{-\alpha}.$$

Since $\alpha > 0$ we have that γ, $\beta < 1$. It follows from Theorem 1 and its corollaries that the number D of discarded intervals is the order of $\epsilon^{-1/n}$. But D is the number N of pieces that comprise the final global approximation, so we have

$$N = \mathcal{O}(\epsilon^{-1/n}).$$

On the other hand, the global error is simply the maximum of the local errors, and each of these is less than ϵ. Thus, we have

$$\| f(x) - A(x) \|_{[0,1],\infty} \leqslant \epsilon.$$

One may eliminate ϵ from these two relations to establish the theorem.

IV. L_p Approximation, $1 \leqslant p < \infty$

We use the norm

$$\| f(x) \|_I = \left[\int_I |f(x)|^p \, dx \right]^{1/p} = \| f(x) \|_{I,p},$$

and consider the same class of functions as before. The assumption about the approximation errors of the local operators is changed slightly to

Assumption 3. *With the notation of Assumption* 2 *we have*

(i) $\text{ERROR}(x, k) \leqslant KF_n(x, k)2^{-k(n+1/p)}$ *if* $[x, x + 2^{-k}]$ *contains no singularity,*

(ii) $\text{ERROR}(x, k) \leqslant K2^{-(\alpha+1/p)}$ *if* $s \in [x, x + 2^{-k}]$.

Recall that $\text{ERROR}(x, k)$ is a bound on $\| (1 - T_I) f \|_{I,p}$. We now determine the rate of convergence of the adaptive method.

Theorem 3. *Assume that* $f(x)$ *satisfies Assumption* 1 *with* $\alpha > -1/p$. *Consider an adaptive algorithm whose local operator satisfies Assumption* 3. *Then the global approximation* $A(x)$ *obtained when the algorithm terminates satisfies*

$$\| f(x) - A(x) \|_{[0,1],p} = \mathcal{O}(1/N^n),$$

where N *is the number of pieces comprising* $A(x)$.

Proof. As in the proof of Theorem 2, there are at most $2R$ distinguished intervals that contain a singularity. The values of γ and β of the partition algorithm are

$$\gamma = 2^{-(n+1/p)}, \qquad \beta = 2^{-(\alpha+1/p)}.$$

Since $\alpha > -1/p$, we have that $\gamma, \beta < 1$. It follows from Theorem 1 and its corollaries that the number N of discarded intervals is the order of

$$N = \mathcal{O}(\epsilon^{-1/(n+1/p)}).$$

To estimate the global error we note that

$$\int_0^1 |f(x) - A(x)|^p \, dx = \sum_I \int_I |(1 - T_I) f|^p \, dx,$$

when the sum is over all the intervals generated when the algorithm terminates. We have from the discard mechanism that

$$\int_I |(1 - T_I) f|^p \, dx \leqslant \epsilon^p,$$

and thus, the global error is bounded from

$$\int_0^1 |f(x) - A(x)|^p \, dx \leqslant N\epsilon^p,$$

which implies

$$\|f(x) - A(x)\|_{[0,1],p} \leqslant N^{1/p}\epsilon.$$

If we eliminate ϵ from this relationship and the one relating N and ϵ, we obtain the conclusion of the theorem and this concludes the proof.

V. Degree of Convergence Results for Spline Approximation

The first degree of convergence results for nonlinear spline approximation [5] involved essentially the same class of functions as defined by Assumption 1. We note that the main theorems of [5] are direct corollaries of Theorems 2 and 3 of this paper using the same local spline operator. The present results do not, however, allow one to establish the more general and much deeper degree of convergence results of Burchard and Hale [3].

There are two advantages of the present approach over the earlier one. First, the present proofs are somewhat simpler (a good deal of the complexity is buried in the theorem used from [7]). More significantly, the present

proof is completely constructive without any a priori knowledge about $f(x)$. The earlier proofs are also constructive but they require that one know both the exact location of all the singularities of $f(x)$ and an upper bound on the strength α of the singularities.

The results of this paper may be used to establish degree of convergence results for other nonlinear piecewise approximation schemes (e.g., using generalized splines of various types) but we have not carried out the details of exhibiting the appropriate local approximation operators.

VI. Local Operators for Smooth Approximation

In an earlier paper [6] we pointed out the need for computational methods for general purpose approximation that

a. are fast to compute,

b. give high accuracy, and

c. give smooth approximations.

At that time, no methods were known with all three of these attributes although methods with any two of them were known. The adaptive methods of this paper allow such computational methods to be constructed by using appropriate local piecewise polynomial approximation operators. We describe three classes of such operators with various values of smoothness and rate of convergence. We let r indicate the smoothness (i.e., $A(x)$ is in $C^{(r)}[0, 1]$) and N (the number of knots) and n (the polynomial degree) govern the convergence properties and computational complexity.

A. *Local Hermite Approximation Operators*

These operators include linear interpolation ($r = 0$, $n = 1$) and cubic Hermite approximation ($r = 1$, $n = 3$). In general we have $n = 2r + 1$ and the polynomial approximation $A_I(f, x)$ on the interval $I = [s, t]$ is determined by

$$\frac{d^k A_I(f, x)}{dx^k} = \frac{d^k f(x)}{dx^k}, \qquad \text{for} \quad x = s, t, \quad k = 0, 1, 2, \ldots, r. \tag{1}$$

If s and/or t are singularities of $f^{(k)}(x)$ then the operator T_I is modified by deleting the conditions involving infinite derivatives and reducing the polynomial degree a corresponding amount. These local operators lead to approximations $A(x) \in C^{(r)}[0, 1]$ with

$$\|f - A(x)\| = \mathcal{O}(1/N^n)$$

where N is the number of knots and n is the polynomial degree.

B. *Local Hermite Approximation Operators with Expanded Degree*

These operators are simple extensions of the preceding ones which increase the polynomial degree while keeping the smoothness fixed. Let $n = 2r + 1 + q$ and then $A_I(f, x)$ on $I = [s, t]$ is determined by the conditions (1) above plus

$$A_I(f, x_j) = f(x_j), \qquad j = 1, 2, \ldots, q, \tag{2}$$

where $s < x_1 \leqslant x_2 \leqslant \cdots \leqslant x_q < t$ (repeated values in the sequence x_j imply interpolation of appropriate derivatives). For example, if $f(x)$ has five derivatives and one only needs $A(x)$ to have continuous slope, then one can keep $r = 1$ and use piecewise quintic polynomials by introducing two interpolation points between the knots.

C. *Local Spline Approximation by Quasi-Interpolants and Moments*

The quasi-interpolant operators introduced by de Boor and Fix [2] are generalizations of the operator introduced earlier by Birkhoff [1]. The quasi-interpolant is of the form

$$A_I(x) = T_I f = \sum_{\pi} (\lambda_j f)\, B_{j,n}(x), \qquad \text{for} \quad x \in I,$$

where π is a partition of [0, 1], $\{B_{j,n}(x)\}_{j=1}^{N+n+2}$ is the B-spline basis for splines on the partition π and λ_j is a certain linear functional. This is a local operator because only $n + 1$ of the $B_{j,n}(x)$ are different from zero on I. A vast variety of linear functionals λ_j may be used and we refer the reader to [2] for an exposition. The nature of the possibilities is illustrated by the following concrete example for cubic splines. Let $\pi = \{t_j = jh\}_{j=0}^{N}$ be an equispaced partition and $I = [t_i, t_{i+1}]$ then

$$A_I(x) = \sum_{j=i-1}^{i+2} [f(t_j) - h^2/6 f''(t_j)]\, B_{i,3}(x).$$

Slightly more complicated formulas exist for nonuniform partitions. One of the main results of [2] is that if $f(x) \in C^{(r)}[0, 1]$ then

$$\|(1 - T_I) f\|_{[0,1],\infty} \leqslant K\omega(f^{(n)}, |\pi|)\, |\pi|^n,$$

where ω is the modulus of continuity and $|\pi|$ is the maximum mesh length. One can easily modify this to conclude that

$$\|(1 - T_I) f\|_{I,\infty} \leqslant K F_r(I)\, |I|^r,$$

where $|I|$ is the maximum length of the interval I and its six neighboring intervals and $F_r(I)$ is the maximum of $F_r(x, k)$ for the same intervals.

The obvious advantage of these operators is that one has $r = n - 1 = 2$. The obvious disadvantage is that the adaptative computations are more complex. This is because

a. Subdividing an interval affects the approximations in the six neighboring intervals. This is because the B-splines $B_{j,3}(x)$ change.

b. The effect of a singularity extends over seven intervals rather than just one. Appropriate modifications must be made in all of them.

c. The value of ERROR(x, k) is more complicated to estimate.

We also note that minor extensions of the proofs must be made to actually cover this case.

REFERENCES

1. G. BIRKHOFF, Local spline approximation by moments, *J. Math. Mech.* **16** (1967), 987–990.
2. C. DE BOOR AND G. J. FIX, Spline approximation by quasi-interpolants, *J. Approximation Theory* **7** (1973), 19–45.
3. H. BURCHARD AND D. F. HALE, Piecewise polynomial approximation on optimal meshes, *J. Approximation Theory* **14** (1975), 128–147.
4. D. S. DODSON, Optimal order approximation by polynomial spline functions, Thesis, Purdue Univ. 1972.
5. J. R. RICE, On the degree of convergence of nonlinear spline approximation, *in* "Approximations with Special Emphasis on Spline Functions," (I. J. Schoenberg, Ed.), pp. 349–365. Academic Press, New York, 1969.
6. J. R. RICE, General purpose curve fitting, *in* "Approximation Theory," (A. Talbot, Ed.), pp. 191–204. Academic Press, New York/London, 1970.
7. J. R. RICE, A metalgorithm for adaptive quadrature, *J. Assoc. Comp. Mach.* **22** (1975), 61–82.